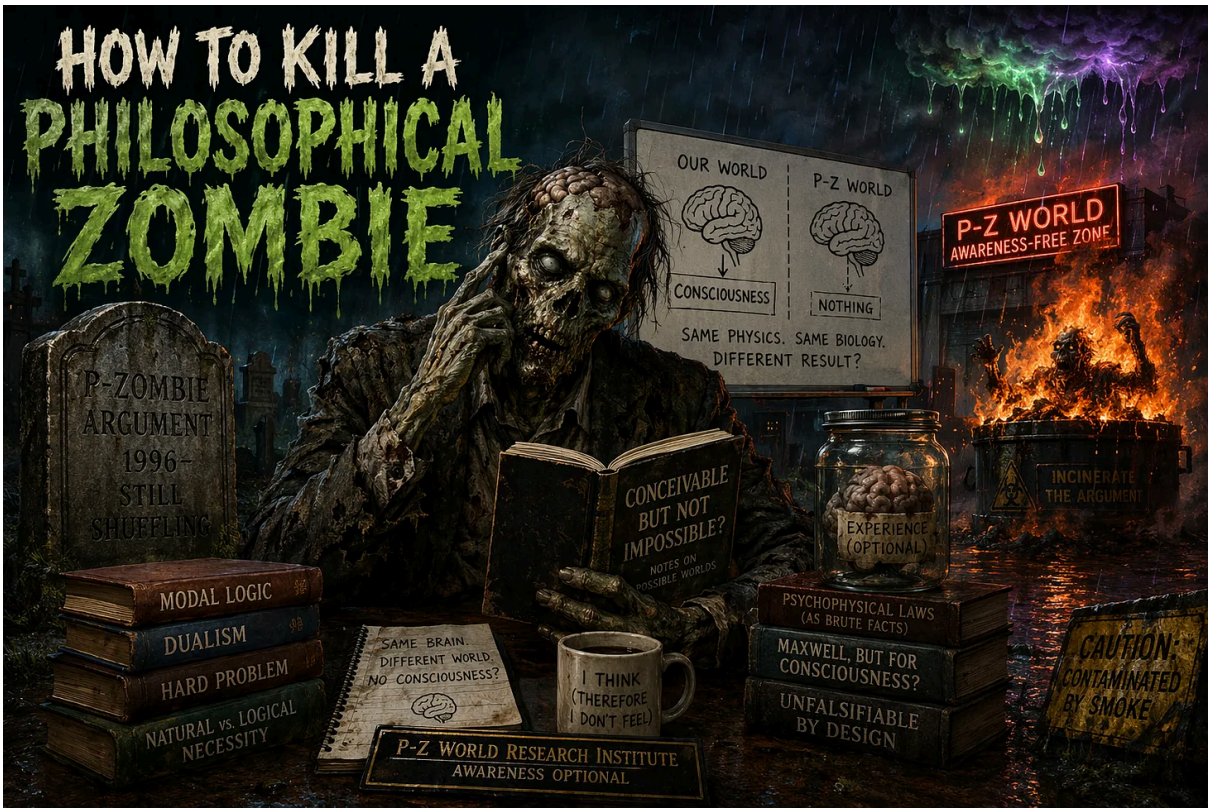
How To Kill A Philosophical Zombie

Advance notice: this essay is a polemic. I didn't mean it to be, but the whole p-zombie thing, I mean, enough already! 28 Years Later and it still dominates philosophy of mind. I had some more criticisms, but these are what made the cut.

I conceive it to be the best zombie essay ever written. My assertion raises no logical contradictions. Therefore there is a possible world where this is the best essay ever written about zombies. Does that mean I definitely wrote a good essay?



A zombie is by definition resistant to dying. They survived the process the first time around, and are now immune to most kinds of attacks. You have to go directly for the brain, or they just keep shuffling along. The p-zombie argument is even more resistant. It isn't entirely alive, but is designed to be unfalsifiable. Its weaknesses are very real, but hidden within layers of deliberately constructed plot armour.

David Chalmers began popularising the philosophical zombie thought experiment back in 1996, but he wasn't its origin. The initial outbreak — patient zero — was Keith Campbell's 'imitation man' in 1970, inverted and released by Robert Kirk in 1974 as an argument in support of dualism. Kirk argued that if a physically identical human can exist without consciousness, then consciousness cannot be physical.

Chalmers didn't initiate the epidemic — his project was a reboot, with a bigger budget and wider distribution.

The key innovation was to give the story a modern sheen using cutting-edge philosophical special effects: modal logic. Special effects work by making the imagined look real — indistinguishable, on screen, from the actual. Modal logic does something similar — it can make what can only be imagined appear to be a genuine possibility.

But no amount of modal complexity can hide this fundamental flaw: the zombie argument only works if you accept in advance that consciousness and physical processes are the kind of things that could be separated. This is not a premise earned — it is a stipulation. A physicalist could assert with equal logical standing that consciousness is an inseparable feature of a functioning human brain. The difference is that the physicalist claim is the one supported by everything we know about brains. Chalmers offers nothing comparable.

Chalmers sometimes retreats to the weaker claim that zombies are at least epistemically possible — that we cannot rule them out, even if they are not metaphysically possible. But epistemic possibility is a measure of ignorance, not of the world's structure.

Before modern chemistry, it was epistemically possible that water was not H_2O. Before modern biology, it was epistemically possible that life required a vital spark, the élan vital, to animate the body. Those possibilities evaporated as knowledge advanced.

The zombie conceivability tracks the same pattern — it marks a gap in our understanding of consciousness, not a gap in nature. It has blurred the line between what can be conceived and what can only be imagined. It is not a genuine metaphysical possibility. It is an imaginary world, not a logically possible one.

There is, of course, no Possible Zombie World out there anywhere. Modal logic was developed as a tool for reasoning about necessity and possibility — a device of logic, not a map of reality. Treating possible worlds as genuine metaphysical locations is a philosophical choice that few people make, not a logical requirement.

In the real world, physical brain structures either produce, interact with, or give rise to consciousness. A hypothetical p-zombie brain would possess those same structures and would be doing the same work, but somehow with no effect. Why they would still be there nobody knows. What those physical structures would be doing if no correlating consciousness exists remains a mystery. How or why the neural structures

that exist in relation to experience would have evolved at all is apparently delegated to science — as if empirical constraints had nothing to do with what is philosophically possible.

But no amount of modal trickery can disguise the fact that an identical brain producing a rich array of consciousness in one instance and a dark empty nothing in another is inconceivable, even if it does not violate the rules of abstract modal logic.

Picture two laptops rolling off the production line. Quality control receives them and turns them on. One loads up the operating system as it should. The other produces nothing, just a blank screen. A tester verifies that the internal processors are operating exactly as they should but something is wrong. One laptop appears faulty. They might open it up to see what is different about the blank one, but they wouldn't find anything — the laptops are physically identical in every way. The only explanation would be that two physical systems can be identical in every way and yet one is simply dark — for absolutely no reason. This is what the zombie argument is asking us to accept.

Suppose that quality control starts swapping components, one by one. They keep going until all of the parts from the functional machine on the left now make up the blank one on the right, and vice versa. Would the machine on the right still be blank? Would the functional one have become blank and the blank one started working? Would there be a moment of transition where one became blank and the other sprang to life? The zombie argument has no answer either way. That is because the scenario, whether concerning laptops or brains, is incoherent.

In physics you might imagine a frictionless surface in order to simplify some calculations. You haven't proven that a frictionless surface could exist anywhere. You have asserted something that you know cannot exist in order to accomplish a specific goal. The zombie argument makes the same move — it stipulates physical identity with no consciousness in advance, and then mistakes that stipulation for a discovery about the nature of mind.

The only escape route is to assert that the fundamental laws of physics and biology operate differently in P-Z World. Chalmers has therefore invented a new kind of fundamental — psychophysical laws — just to keep the zombie going. He argues that the fundamental laws of nature, the irreducible brute facts about the universe, could be different in a possible world. This is how far he must retreat from everything we know of our world in order to keep his theory alive.

Alternative psychophysical laws are imagined in order to protect the zombie argument from the accusation that it is unscientific. If biology and physics operate differently in P-Z World then p-zombies do not contradict science — they are in fact exactly what you would expect the fundamental laws of P-Z World to produce. But this is not empirical science as we know it. It is a fiction, arising fully formed through the act of defining a p-zombie. It only operates if you accept the premise in advance. It is a rationally constructed make-believe version of science rather than empirical science.

Psychophysical laws are not an explanation. They are a shield. A theory that requires identical matter to behave completely differently in an alternative world, for no stated reason, has not solved the problem. It has just made itself unassailable by design — the very definition of plot armour.

I might propose the existence of a pixie operating outside of time and space, magically creating inner experience, and matching it up with our brains on the spot. Science cannot refute my claim, and I assert that the pixie is not governed by our laws of nature but that it creates its own psychophysical laws. Would I really be asserting a philosophical position though?

Relying on unknown and unimaginable fundamental laws is little more than a philosophical convenience that gives a dead theory the appearance of life. It is the same kind of circularity that runs through the entire project: assume psychophysical laws that support the existence of an awareness-free brain in order to reveal that consciousness is distinct from the brain.

Chalmers attempts to support these fictional laws by analogy. When electromagnetic phenomena couldn't be explained by Newtonian mechanics, Maxwell didn't try to reduce electromagnetism to mechanics — he posited new fundamental laws. Chalmers says consciousness requires the same move. He's wrong.

Maxwell posited electromagnetic fields because the evidence demanded them — there were measurable phenomena that couldn't be explained without them. Chalmers is positing psychophysical laws to satisfy an intuition: that because awareness doesn't feel like the same kind of thing as atoms and molecules, it cannot possibly be explained in purely physical terms.

Maxwell revised physics to match the evidence. Chalmers revised it to match a gut-feeling. They are not comparable.

Physical laws constrain how systems behave. Gravity constrains what matter does. Electromagnetism constrains how charged particles interact. Change the physical structure and the outcomes change accordingly. That is what a law is — it arises from

the structure it governs, constrains what that structure can do, and supports counterfactual reasoning about what would happen if things were different.

The proposed psychophysical laws do none of this. How could they when they are invented from nothing?

They do not arise from the physical structure since the same structure can apparently produce consciousness or not, depending on which world you happen to be in. They do not constrain physical behaviour — everything physical remains identical regardless of whether consciousness is present or absent. They do not support counterfactual variation — you cannot change the physical facts to get a different experiential outcome, because the physical facts are stipulated to be identical either way.

All they do is declare that in one world consciousness appears and in another it does not, while leaving every physical fact untouched. If an appeal to psychophysical laws makes a p-zombie conceivable — almost anything is conceivable, including transcendental pixies.

A philosophical theory that ignores empirical facts ought to be considered weaker as a result, not more unassailable. The brain is the product of millions of years of evolution. Every feature developed in response to specific survival conditions. Evolution exerts enormous pressure against large-scale non-functional complexity. The brain's metabolic cost is vast. In our world, that cost is justified by what the brain does.

P-Z World asks us to accept an identical structure that evolved for entirely different reasons — or perhaps no reason at all. That is not a possible world. It is an incoherent one.

Humans have an extended frontal cortex, one that has specifically evolved to generate or interact with consciousness. The zombie world isn't just a world that happens for no reason to have different psychophysical laws. It is one that has produced an uncanny coincidence — an identical brain, the most complex and multi-functional object in the known universe, but for totally different reasons and doing totally different things.

The p-zombie hypothesis is not merely science-removed. It is scientifically incoherent. It creates an explanatory gap that makes the hard problem look like a hairline fracture. Evolution in the real world has taken millions of years of fine-tuning to produce brains with rich inner lives and an ability to discuss and reflect on them.

No ‘conceivable’ evolutionary path could account for this convergence. P-Z World is more belief than philosophy — unfalsifiable by design and untouchable by evidence.

The entire apparatus — complex modal logic, epistemic conceivability, inexplicable psychophysical laws — all of it constructed to protect a single intuition: that experience cannot be purely physical. Chalmers acknowledges it cannot be demonstrated empirically or shown to be logically necessary. The structure is unassailable all the same, not because the argument is sound, but because no objection can fully land on a position that is intricately layered to be unfalsifiable.

How do you kill a philosophical zombie? It turns out there is no way. This is not a Romero film, where the rules hold and the headshot does its job. This is Return of the Living Dead — the reboot where the zombies can never die.

In that film they eventually try incineration. They think the zombies are finally gone. Then the rain comes, contaminated by the smoke from the burning corpse, and the whole thing starts again.

In fiction everything is negotiable. The rules of reality don’t apply, and writers can always find a way to keep a successful franchise going. So how do you kill a philosophical zombie? You cannot kill what is already dead.

About the Author

Sopa Thaye is a secondary school Philosophy and Religion teacher in the UK and author of the philosophy Substack *The Primacy of Perception*. He is currently completing a book on epistemology, consciousness, and metaphysics.

Publication Information

Originally published April 2026 at:
https://sopathaye.substack.com/p/how-to-kill-a-philosophical-zombie

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.